1-1-2017

Determination Of The Direct Protein-Protein Interactions In The Drosophila Sin3a Complex

Ian Moore

Wayne State University,

DETERMINATION OF THE DIRECT PROTEIN-PROTEIN INTERACTIONS IN THE *DROSOPHILA* SIN3A COMPLEX

By

IAN MICHAEL MOORE

THESIS

Submitted to the Graduate School

of Wayne State University,

Detroit, MI

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

2017

MAJOR: BIOLOGICAL SCIENCES

Approved By:

_____

Advisor Date

DEDICATION

To my parents, siblings, friends and lovely fiancé:
I would not be where I am without your support.

ACKNOWELDGEMENTS

I would first like to thank my advisor Lori Pile for everything she has done for me over this past two and a half years. I have grown immensely both personally and professionally, and have learned to think and speak like a true scientist. But more than that, you have been the best mentor I could have asked for. I want to thank Valerie Barnes for being there to provide input, talk things through and to handle everyday situations in the lab. I would also like to thank all current and former lab members, as you have made every day in the lab an enjoyable experience. You are much more than my peers and colleagues; you are my friends. For that I am forever grateful.

I would also like to thank my committee members, Drs. Athar Ansari and Penelope Higgs. From teaching my courses to providing protocols, discussing results and lending me reagents, I would never have made the progress I did without your guidance and assistance. I cannot thank you enough for all you have done. To all the staff here at Wayne State from professors to the front office: thank you for all the help you have provided.

I want to thank Patrick McLaughlin. We came into this program together, and I can honestly say that I would not have had nearly as enjoyable an experience were it not for you being across the hall from me.

Finally, I want to thank my parents, siblings, future in-laws and fiancé. You have all been more encouraging that I could ever dream of during this stage of my life. From checking in to make sure I was doing ok, to asking questions about my research, to just providing me a getaway from the lab. Thank you, and I love you.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1 Introduction

Mechanisms and Role of Genetic Regulation

The regulation of genetic material is integral to the health of both individual cells and whole organisms. This regulation allows for the tight control of finely tuned cellular processes such as cell cycle progression, transcription and apoptosis (Aregger and Cowling, 2016; Chen et al., 2013; Gan et al., 2016; Liang et al., 2012). Both proteins and DNA regulatory elements play a role in maintaining the integrity of the cellular processes, allowing for maintenance of the health of the cell given the correct intra- and extracellular environment while reserving the ability to promote cell death when given the appropriate signals.

Compromising this regulation of genetic material can result in the onset of many diseases, such as cancer, autoimmune diseases and neurological disorders (Lee and Young, 2013; Villard, 2004; Wang et al., 2016; Wen et al., 2016). As genetic diseases are usually chronic and have long lasting effects on the affected organism, elucidating the mechanisms that influence expression and silencing of genes are of great interest to researchers. One mechanism that has been demonstrated to play a role in gene expression is the addition and removal of chemical groups to histone tails (Bannister and Kouzarides, 2011; Berger, 2002). This process can be classified as a post-translational modification as the chemicals are added or removed from the histone following the translation of the mRNA coding for the protein. The addition and removal of chemical groups is thus dynamic and varies throughout the cell cycle and in response to different environmental cues (Feil and Fraga, 2012).

Nucleosomes are made up of histone octamers around which DNA is wrapped to promote compaction of genetic material, facilitating the fit of DNA into the nucleus of a cell. This histone-DNA complex is referred to as a nucleosome. The histones include an N-terminal tail, which can be subjected to the addition and removal of chemical groups. These marks can serve as docking ports for proteins involved in transcription or silencing, or can alter the charge of the histone tail, which can in turn repel the negatively charged DNA (Kunowska et al., 2015; Lee et al., 1993; Rossetto et al., 2012). The repulsion of the DNA from the histone tail results in more accessible DNA, which can then be bound by different factors involved in the regulation of transcription (Zhang et al., 2016; Zhang et al., 2012).

Two specific chemical groups that can influence transcription when placed on histone tails are acetyl groups and methyl groups. Acetyl marks neutralize the positive charge of the histone tail, which repels the negatively charged DNA from the histone, resulting in a more open structure (Lee et al., 1993). Acetyl groups on histones can also be recognized by bromodomains of proteins, which leads to a “docking” of the protein on the histone (Dhalluin et al., 1999). Methyl groups can also be used as docking points for proteins that lead to an increase or decrease in transcription. One protein known to bind to methylated histones is heterochromatin protein 1 (HP1), which can multimerize with other HP1 proteins on nearby nucleosomes to compact chromatin and limit accessibility (Hiragami-Hamada et al., 2016; Mishima et al., 2015).

As the chemical identity of groups placed on histone tails can play such a large role in gene regulation and consequently the health of an organism, there is great interest in understanding the mechanism through which these groups are added. To fully

understand a molecular mechanism of action, it is necessary to understand the proteins involved. Proteins that either add or remove the chemical marks on histone tails are known as histone modifying enzymes. As they can have a direct influence on the expression of key regulatory genes, these enzymes and their associated proteins are a major focus of research. Results of such research can have implications for treatment of genetic diseases.

SIN3 Isoforms and Associated Proteins

One protein complex that is associated with both a histone deacetylase (HDAC) and a histone demethylase (KDM) is the SIN3 complex. SIN3 is a protein that is conserved from yeast to mammals and is a key regulator of the genome (Kadamb et al., 2013).

In *Drosophila*, a single *Sin3A* gene encodes three different SIN3 isoforms (Neufeld et al., 1998; Pennetta and Pauli, 1998). These three isoforms are known as SIN3 220, SIN3 190 and SIN3 187 and are named for their molecular weight in kilodaltons. These isoforms are identical throughout much of their structure including the presence of four paired amphipathic helix (PAH) domains and a histone deacetylase interaction domain (HID) (Fig 1) These domains are conserved from yeast to mammals (Neufield et al. 1998; Pennetta and Pauli 1998). PAH domains consist of alpha helices, which mediate interactions with other proteins. The HID mediates interaction with the histone deacetylase RPD3. The only difference between the SIN3 isoforms is the inclusion of unique exons at their C-terminus as the 220, 190 and 187 isoforms have a unique stretch of 315 amino acids, 31 amino acids and 5 amino acids, respectively. The two major isoforms present in *Drosophila* are the 220 and 187 isoforms (Sharma et al., 2008).

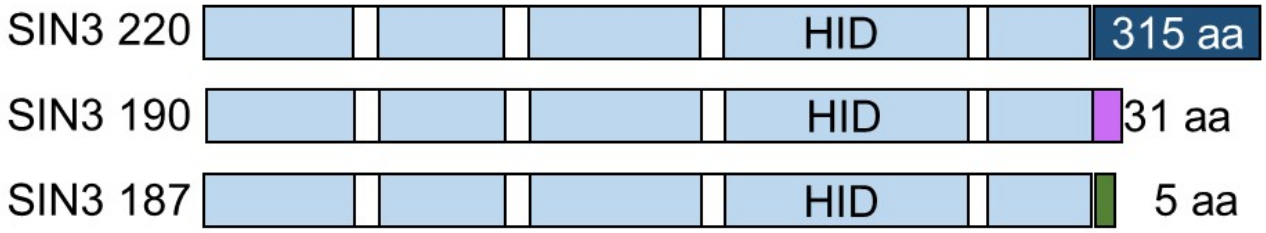

Figure 1. *Drosophila* SIN3 isoforms. Three *Drosophila* isoforms are generated through alternative splicing from one *Sin3A* gene and are named after their molecular weight in kilodaltons. White regions of the structure represent paired amphipathic helix (PAH) domains, while the region labelled HID represents the histone deacetylase interaction domain. Unique exons are indicated by the blue, purple and green boxes at the C-terminus of each protein.

Both SIN3 220 and SIN3 187 are capable of forming complexes with a core group of proteins, which includes ING1, Sap130, SDS3, RPD3, BRMS1, Pf1 and Arid4B (Spain et al., 2010) (Fig 2). The SIN3 220 complex, however, also has associations with the proteins LID, EMSY and p55 (Spain et al., 2010). RPD3 is the homolog of mammalian HDAC1, which is a histone deacetylase. LID is the homologue of mammalian KDM5, which acts as a histone demethylase. Both of these proteins are involved in the removal of chemical groups on histone tails, enabling the regulation of gene expression.

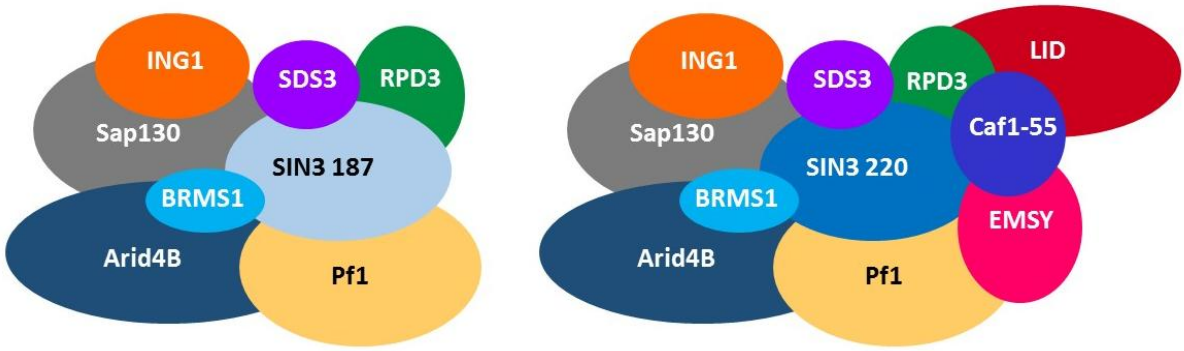

Figure 2. *Drosophila* SIN3 complexes. SIN3 187 and SIN3 220 are capable of forming complexes with a core group of proteins, with SIN3 220 interacting additionally with Caf1-55, LID and EMSY

Adapted from Spain et al. 2010

The Role of SIN3

In both *Drosophila* and mice, knockout of *SIN3A* results in embryonic lethality (Cowley et al., 2005; Dannenberg et al., 2005; Neufeld et al., 1998; Pennetta and Pauli, 1998). Studies using *Drosophila* polytene chromosomes, microarrays, RNA-seq and ChIP-seq have shown that SIN3 binds throughout the genome and regulates between 2-12% of the genome, depending on the assay used (Gajan et al., 2016; Pile et al., 2002; Pile and Wassarman, 2000; Saha et al., 2016). Additionally, knockdown of *Sin3A* in *Drosophila* S2 cells results in cellular proliferation defects and an arrest in the G2/M phase (Pile et al., 2002; Swaminathan and Pile, 2010). Together, these data show that SIN3 is an essential protein with the ability to regulate genes vital for development and survival.

Experiments have also been performed to look at both the expression of the 220 and 187 isoforms as well as the processes that they regulate. Previous work has determined that there is a turnover between the two isoforms throughout development. The SIN3 220 isoform is highly expressed early in development and in proliferative cells, such as larval imaginal wing discs (Sharma et al., 2008). As cells begin to exit their proliferative phase, however, there is an increase in the 187 isoform as well as a subsequent decrease in the level of the 220 isoform (Chaubal et al., 2016; Sharma et al., 2008).

Studies have also identified subsets of genes regulated by SIN3 220 and SIN3 187 (Saha et al., 2016). This work identified similar and distinct binding sites of both isoforms across the *Drosophila* genome. These binding site data sets were then combined with RNA-seq data. It was determined that there are unique and common gene sets directly regulated by the 187 and 220 isoforms (Saha et al., 2016). Both isoforms have been shown to regulate the processes of metabolism, cell cycle and morphogenesis and neuron development, while the 187 isoform exclusively regulates the additional processes of endocytosis, apoptosis, phosphate metabolism and post-embryonic development (Saha et al., 2016).

In addition to revealing the overlapping and distinct roles of the 220 and 187 isoforms as well as their associated proteins, studies have revealed interesting data concerning the ability of the isoforms to perform compensatory functions. Studies in *Drosophila* have shown that the lethality in *Sin3A* null mutants can be rescued through expression of ectopic SIN3 220, but not through the expression of the 187 isoform (Spain

et al., 2010). These data combined with the overlapping and distinct gene sets regulated by the differing isoforms enforces the idea that these two similar yet distinct isoforms have isoform-specific roles in the development and life cycle of *Drosophila*.

SIN3 and Disease

As SIN3 has the ability to regulate a substantial proportion of the genome, it is not surprising that the SIN3 complex has been demonstrated to regulate genes involved with disease. Specifically, SIN3 has been implicated in diseases such as cancer, through inappropriate recruitment to genes and their subsequent silencing (Farias et al., 2010; Kwon et al., 2015), and Alzheimer's, where it was shown in *Drosophila* that mutations can result in enhancement of the Alzheimer's β-amyloid phenotype(Cao et al., 2008).

The SIN3 complex has been shown to negatively regulate a number of oncogenes and stabilize proteins associated with normal growth and development of the cell. As described in previous research, SIN3 stabilizes the tumor suppressor p53, protecting it from degradation (Zilfou et al., 2001). SIN3 has also been identified as a negative regulator of *Myc*, *E2F*, and *Rb* (Dannenberg et al., 2005; Garcia-Sanz et al., 2014).

Members of the SIN3 complex are also implicated in cancer progression. Pf1, which is common to both the 220 and 187 isoform complexes, is overexpressed in breast cancer (Bansal et al., 2015). It works to recruit EMSY and JARID1B, which are both factors in breast cancer, in cancerous cells (Kwon et al., 2015). Disruption of the Pf1-SIN3 interaction using peptides or knockdown of Pf1 in triple negative breast cells showed a reduced ability of tumor cells to form colonies (Bansal et al., 2015). This is in part because Pf1 is shown to interact with the transcription factors NANOG, OCT4, and SOX2

(Bansal et al., 2015), which are hallmark proteins of tumor stem cells, which are undifferentiated proliferative cells seen in many cancers (Weissman, 2015).

Studies have also been performed on Inhibitor of Growth 1, or ING1, a common component of the 220 and 187 isoform complexes. Reduced levels of ING1 can result in breast cancer (Thakur et al., 2014). Interestingly, metastatic cells that were injected into mice and made to overexpress ING1 exhibited a slower rate of growth than those cells that did not ectopically express ING1 (Thakur et al., 2014). Coupled with the roles that the SIN3 complexes play in genome-wide regulation and the association with histone-modifying enzymes, the study of SIN3 complex proteins and their mechanisms of action present an attractive target for the treatment of genetic diseases.

Importance of Protein Interactions

One key component to the identification and understanding of the mechanism by which a protein or protein complex acts on a substrate is knowing the proteins with which is associates (Schwikowski et al., 2000). Protein interactions can help tease apart cellular processes and regulations (Jackson et al., 1991; Letovsky and Kasif, 2003). Additionally, protein interactions can play key roles in the stabilization of protein complexes (Gibson, 2009; Perkins et al., 2010). Without this stabilization, the complex may not be able to perform its intended function, which can result in negative consequences for the cell (Sato et al., 2000; Sherman et al., 2001).

Interestingly, previous work has shown the necessity of almost all proteins involved in the SIN3 complex. Knockdown of all components of the SIN3 complex with the exception of EMSY results in embryonic lethality in *Drosophila* (Laity et al., unpublished).

These data confer additional evidence of the importance of both the SIN3 isoforms themselves as well as the proteins with which they associate.

Previous research has identified direct interacting partners within the SIN3 complex. Some interactions include SIN3 and RPD3, SIN3 and Pf1 (Yochum and Ayer, 2001), Caf1-55 and SIN3 (Saha, 2017), Caf1-55 and RPD3 (Marhold et al., 2004) and ING1 and Sap130 (Giot et al., 2003). Not all studies, however, have looked exclusively at direct interactions and instead use high throughput methods to identify complexes as a whole. Although there is evidence of direct interaction between some proteins associated with the SIN3 complex, there have been no assays performed to identify all interactions within the *Drosophila* SIN3 complex. Additionally, limited work has been done to distinguish the regions of SIN3 or associated proteins that are responsible for the recruitment of the SIN3 220 complex specific proteins LID, EMSY and p55.

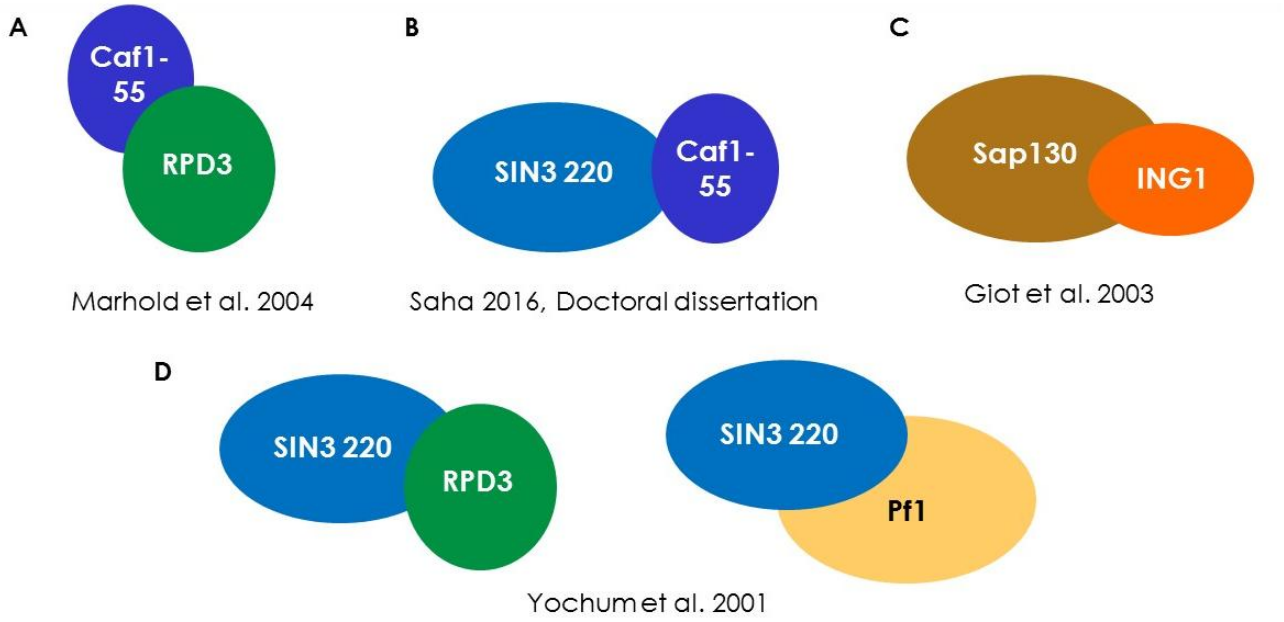

Figure 3. Previously identified direct interactions between *Drosophila* SIN3 complex proteins.

As the SIN3 220 and 187 isoforms have similar yet distinct roles in the life cycle of *Drosophila*, it is possible that the interactions of the SIN3 220 complex with p55, LID and EMSY may play a role in driving the specificity of the 220 complex in the *Drosophila* life cycle. Therefore, it is of great interest to determine the direct interactions occurring with SIN3 220, specifically with the unique stretch C-terminal stretch of 315 amino acids.

Project Summary

The goal of this project is to determine the direct interacting partners within the *Drosophila* SIN3 complex. A bacterial expression system was used to perform immunoprecipitations to assay physical interactions *in vitro*. Proteins that interact directly with the unique 315 amino acids of SIN3 220 were identified during this research, as well

as other interactions within the complex. *In vivo* assays were also performed in S2 cells to determine the region of p55 responsible for maintaining the interaction with SIN3 220. Immunoprecipitations were performed in S2 cells containing a truncated form of p55. A region of p55 required for the interaction with SIN3 was identified. Taken together, these studies lend insight into the interactions of proteins within the SIN3 complex. This knowledge will provide potential new targets for experiments that aim to correlate the loss of specific SIN3 complex components and the ability of the complex to regulate the genome.

Chapter 2 Materials and Methods

Creation of Bacterial Expression Clones

The pETLX vector was created to introduce additional unique restriction sites into the multiple cloning site (MCS) of a bacterial expression vector for ease of cloning. To create the pETLX vector, the pET28(a) vector was digested with the *EcoRI* and *XhoI* enzymes (NEB), which each cut in one spot in the multiple cloning site. Complementary primers were designed with a sequence that contained *EcoRI*, *NotI*, *AvrII*, *Acc651*, *MfeI*, *ScaI* and *XhoI* sites. The primers were then annealed and digested using *EcoRI* and *XhoI.* The fragment was then ligated into the digested pET28(a) vector.

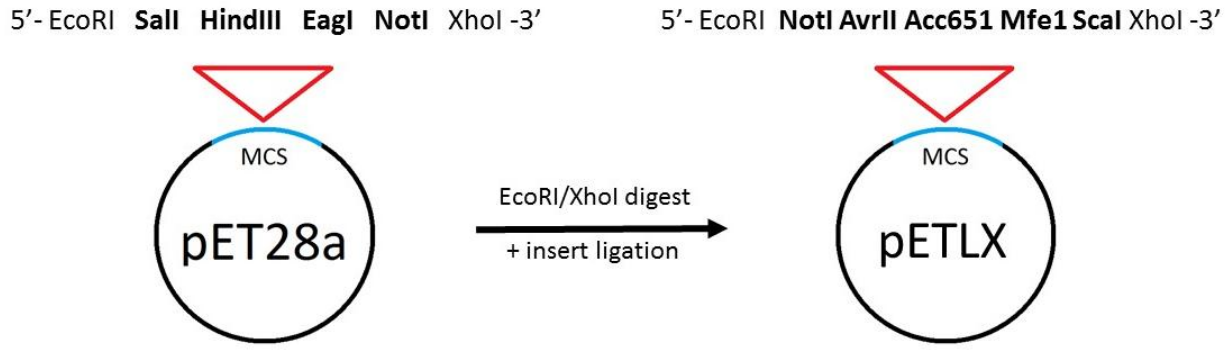

Figure 4. pETLX creation. Complementary primers were designed containing the sequence for *EcoRI*, *NotI*, *AvrII*, *Acc651*, *MfeI*, *ScaI* and *XhoI* restriction sites. The pET28a plasmid was then digested with *EcoRI* and *XhoI* and the backbone isolated. The complementary primers were annealed and ligated into the pET28a backbone, resulting in pETLX.

To generate the clones for expression of SIN3 complex proteins, I amplified genetic sequences using either BDGP orfeome clones or a cDNA library prepared from S2 cell RNA. BDGP orfeome clones were provided by the *Drosophila* Genomics

Resource Center (DGRC) and contain sequences that encode *Drosophila* proteins that can be expressed in S2 cell culture. Primers used during amplification contained restriction sites that would allow the ligation of the gene into the appropriate vector. Clones were confirmed through restriction digestion analysis and sequencing. The sequences of primers used for cloning and the restriction sites used are found in Table 1.

Table 1. Primers/Restriction sites used for cloning into bacterial expression vectors

Primer	Primer Sequence	Restriction site
BRMS1BamHIFor	CGTATAGGATCCATGCCCGTGAAGAATGGC	BamHI
BRMS1XhoIRev	CGATATCTCGAGCTATCCGCTAGCTCCTGC	XhoI
EMSYEcoRIFor	GCATATGAATTCATGTGGCCGCAAACGCTGG	EcoRI
EMSYAvrIIRev	GCTGCGCCTAGGTTAGGGTAACTTAATTGCT	AvrII
ING1EcoRIFor	CGTGTAGAATTCATGATAAACCCCATTGCATCC	EcoRI
ING1AvrIIRev	CGTATACTCGAGCTAGGTCTTCTCCTCCTTTTCC	AvrII
Sap130NotIFor	GAAGTTATCAGTCGACGCGGCCGCTATGAGTGCGCCTAGTG	NotI
Sap130Acc651Rev	GGTCTAGAAAGCTTGCCTAGGTACCGATCTTTTCGCGC	Acc651
SDS3EcoRIFor	GCACGTGAATTCATGTCCAACTACTACAGTCTACTAC	EcoRI
SDS3XhoIRev	CAATATCTCGAGCTATCGGCGCTTTATCGAGATC	XhoI

Induction of Protein via IPTG and Determination of Solubility

The bacterial clones of interest, in *E. coli* BL21(DE3) cells, were induced using the IPTG induction system, which has been previously described (Volynets et al., 2016). In short, bacterial colonies harboring plasmids of interest were grown in LB and the appropriate antibiotic. A larger volume of LB plus antibiotic was then inoculated with this culture and the A_{550} was measured until the density of the culture reached midlog (0.5-

0.7). The culture was then induced with a final concentration of 1 mM IPTG. After induction, samples of culture were taken at different time points and subjected to centrifugation. The pelleted samples were then resuspended in Laemmlli Buffer (Bio-Rad) and the lysates subjected to SDS-PAGE. The gel was then stained with Coomassie blue to visualize any expression of the recombinant protein.

A subsequent culture was then subjected to the conditions under which protein expression was observed. The culture was then centrifuged at 3220 x g for 15 minutes at 4^{o}C. Lysis of bacterial cells was carried out as previously described (Saha, 2017). Briefly, cells were resuspended in bacterial lysis buffer (500mM NaCl, 0.1% lysozyme (MP Biopharmaceuticals), 0.001% universal nuclease (Pierce) and protease inhibitor tablet (Roche) (1 tablet/10ml) in 0.9X phosphate buffer saline (PBS)) and left on ice for one hour. Samples of whole cell lysates were stored, and the remaining cells were sonicated at 30% amplitude for two minutes total of 0.5 seconds on and 0.5 seconds off. After sonication the culture was subject to centrifugation at 4^{o}C for 30 minutes at 3220 x g. The supernatant was then stored as the soluble fraction.

Whole cell lysates and soluble fractions were then subjected to SDS-PAGE and transferred to a PVDF membrane. The PVDF membrane was then subjected to a Western blotting protocol. The relative amount of protein in the soluble fraction compared to the insoluble was then determined by visual inspection of developed western films.

Induction of Protein via Autoinduction and Determination of Solubility

The bacterial clones of interest in *E. coli* BL21(DE3) cells were induced using the autoinduction protocol as previously described (Studier, 2005). Briefly, bacterial colonies harboring plasmids of interest were grown in LB and the appropriate antibiotic. This culture was then used to inoculate a culture of autoinduction media (1 mM $MgCl_2$, 0.01% glycerol, 0.001% glucose, 0.004% lactose, 25 mM mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4 in LB). These cultures were grown in a 37^{o}C shaker for 6-8 hours before being placed in a 20C shaker overnight. Following autoinduction, cultures were centrifuged and resuspended in bacterial lysis buffer as previously described (Saha, 2017). After cells were resuspended and left on ice for one hour, a sample of the lysate was saved as the whole cell lysate. The remaining sample was sonicated as described above and subjected to centrifugation at 4^{o}C for 30 min at 600 x g. The supernatant that resulted from the centrifugation was saved as the soluble fraction. A volume of Laemmli sample buffer equivalent to the original amount of lysate subjected to the 600 x g centrifugation was used to resuspend the remaining aggregate. This sample was saved as the pellet sample.

An SDS-PAGE gel was then loaded with the whole cell lysate, soluble and pellet fractions. The samples were then subjected to SDS-PAGE and transferred to a PVDF membrane. The membrane was then subjected to a Western blotting protocol. The relative amount of protein in the soluble fraction compared to the insoluble was then determined by visual inspection of developed western films.

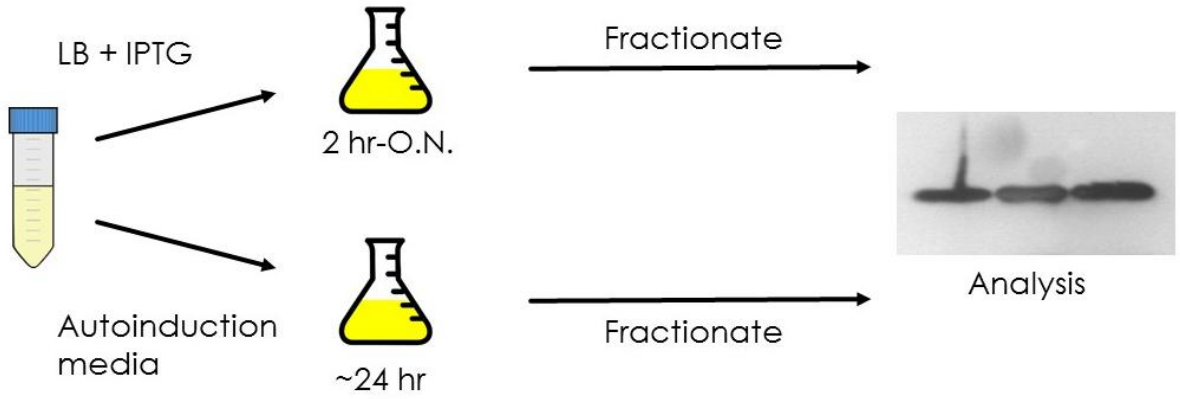

Figure 5. Schematic of IPTG and autoinduction strategies. An aliquot of overnight culture is used to inoculate autoinduction media (top) or LB + IPTG (bottom). Cells are fractionated at approximately 24 hours (autoinduction protocol) or after a pre-determined amount of time post-IPTG induction. The fractions are then subjected to SDS-PAGE and visualized by Western blot.

Prediction of Protein Structure

The sequence of the unique C-terminus of the SIN3 220 isoform was analyzed using software to predict the presence of ordered and disordered regions. The sequence corresponding to the unique stretch of 315 amino acids was obtained through Flybase and was put into the IUPRED and ANCHOR software.

Immunoprecipitation and Determination of Interactions with the SIN3 220 Unique C-terminus

Plasmids encoding HA and maltose binding protein (MBP)-tagged unique SIN3 220 C-terminal (Uni-C) amino acids and HIS-tagged Caf1-55 were transformed together into *E. coli* BL21(DE3) cells. The HA-tagged SIN3 220 Uni-C and HIS-tagged ING1 plasmids were also transformed together into *E. coli* BL21 cells. The HA-tagged SIN3 220

Uni-C, HIS-tagged SDS3, HIS-tagged EMSY, HIS-tagged RPD3 and HIS-tagged Sap130 were each transformed into *E. coli* BL21(DE3) cells separately. For all interaction experiments, BL21(DE3) cells were induced using the above described autoinduction method.

For experiments with SIN3 220 Uni-C and Caf1-55 or SIN3 220 Uni-C and ING1, the autoinduced bacterial culture was fractionated as described above. The soluble fraction was then used for immunoprecipitation. Immunoprecipitation of the soluble fraction was performed as previously described (Saha, 2017). Briefly, the soluble fraction was incubated with anti-HA resin at 4^{o}C overnight. The resins were then washed and the bound fraction was eluted using Laemmli sample buffer. Western blotting was used to analyze the bound fraction.

Cultures harboring individual expression plasmids were induced using the autoinduction method and fractionated as described above. The soluble fraction containing the HA-tagged SIN3 220 Uni-C was incubated with anti-HA resin for one hour at 4^{o}C. The resin was then washed and the soluble fraction containing either SDS3, RPD3, Sap130 or EMSY was then added to the anti-HA resin already bound to SIN3 220 Uni-C and incubated at 4^{o}C overnight. The resin was then washed and the bound fraction eluted using Laemmli sample buffer. Western blotting was used to analyze the bound fraction.

Negative control experiments were conducted by incubating anti-HA resin with lysate only containing HIS-tagged complex proteins at 4^{o}C overnight. Immunoprecipitation was performed using the previously described protocol (Saha, 2017).

Creation of Caf1-55med Construct and Cell Line

To determine amino acids integral for the interaction between Caf1-55 and SIN3 220, a truncated mutant of the Caf1-55 was created. This mutant was called “Caf1-55med” (Anderson et al., 2011). The first through 303rd amino acids of Caf1-55 were amplified by PCR from an expression clone containing Caf1-55 in the pMK33 vector (FMO025549), which was purchased from the Berkeley *Drosophila* Genome Project (BDGP). The amplified Caf1-55 construct had a stop codon replacing the 304th amino acid, tryptophan. The FMO025549 plasmid containing the full-length Caf1-55 was then digested with *BamHI* and *XbaI* to release the full-length Caf1-55, and the plasmid backbone was purified. The PCR product of nucleotides encoding the first 303 codons of Caf1-55 was digested with *BamHI* and *XbaI* and ligated into the empty pMK33 vector and this new construct was transformed into DH5α cells. Colonies were used to inoculate LB with ampicillin and the culture was incubated overnight. Plasmid DNA was isolated the next day. The clone containing the Caf1-55med construct was confirmed through PCR and restriction digestion. The plasmid was sent to Eton Biosciences for sequencing. The primers used for amplification of the Caf1-55med sequence can be found in Table 2.

Table 2. Primers/Restriction sites used for cloning into S2 expression vectors

Primer	Sequence	Restriction Site
Caf1-55medFor	CGCGGGATCCATAACTTCG	BamHI
Caf1-55medRev	GCGCTCTAGATAGAGCTACAGTC	XbaI

The creation of Caf1-55med resulted in a truncated form of Caf1-55 that contained a C-terminal FLAG and HA tag. A stable transgenic S2 cell line for the expression of Caf1-55med was generated using the protocol previously described (Gajan et al., 2016).

Briefly, the Caf1-55med construct was transfected into *Drosophila* S2 cells cultured in *Drosophila* Schneider's media (1X) + L-glutamine with 10% heat-inactivated fetal bovine serum (Invitrogen) with gentamycin added to a final concentration of 50 mg/ml. Following selection of transfected cells, the stable S2 cells containing Caf1-55med were grown in media containing 0.1 mg/ml hygromycin. Cells were kept at 27ºC.

Caf1-55med Nuclear Fraction Expression and Immunoprecipitation

Nuclear fractions were prepared from S2 or stable transfected cell lines expressing wildtype Caf1-55 or Caf1-55med were performed using the method that has been previously described (Spain et al., 2010). 850 µl of nuclear extract was incubated overnight at 4ºC with 150 µl of interaction buffer (20 mM HEPES (pH 7.4), 150 mM NaCl, 0.5 mM EDTA, 1% Triton X-100) and 35 µl of anti-HA resin (Sigma). The resin was subsequently subjected to five minute washes at 4ºC with wash buffer 1 (150 mM NaCl, 0.1% sodium deoxycholate, 0.1% SDS, 1% Triton X-100, 10% glycerol and 20 mM Tris (pH 7.4)), wash buffer 2 (500 mM NaCl, 1.5% Triton X-100, 0.1% sodium deoxycholate, 0.5 mM EDTA, 10% glycerol and 20 mM HEPES (pH 7.4)) and wash buffer 3 (300 mM NaCl, 0.1 mM EDTA, 10% glycerol, 1.5% Triton X-100, 1 mM $MgCl_2$, and 20 mM HEPES (pH 7.4)). The bound fraction then eluted by the addition of 35 µl of Laemmli Buffer (Bio-Rad) to the resin and incubating at room temperature for 30 minutes. Expression of the tagged proteins was verified and the bound fractions from immunoprecipitation assays were analyzed by Western blot analysis.

Western Blot

All Western blotting was performed according to established protocols (Sambrook J, 2001). Samples subjected to SDS-PAGE was transferred to membrane using a wet

transfer. Blots were probed with the following primary antibodies: SIN3 220 (1:2000 (Pile and Wassarman, 2000)), HIS (1:1000; Qiagen), Caf1-55 (1:3000; Abcam) and HA-HRP (1:7000; Sigma). Donkey anti-rabbit HRP conjugated IgG (1:3000; GE Healthcare) and sheep anti-mouse HRP conjugated IgG (1:3000; GE Healthcare) were the secondary antibodies used. Signals on blots were detected using ECl or ECl Prime (GE Healthcare) and exposure to autoradiography film (Alkali Scientific).

Chapter 3 Analysis of Possible Interactions of SIN3 Complex Components

3.1 Introduction

To fully appreciate and understand the mechanism of a biological or chemical process in the cell, it is important to know the context in which this process occurs. Protein interactions are examples of context that could be important for deciphering a mechanism of action. Interactions between proteins with distinct enzymatic activities can be required to catalyze a process or interactions may be necessary to confer a conformational change. In turn, this conformation change gives the protein an intrinsic ability to act upon a substrate.

To gain a better insight into what drives the differences between the SIN3 187 and 220 complexes, the unique C-terminal region of the SIN2 220 isoform was used for immunoprecipitation with other complex components. Proteins were prepared using the bacterial expression system. The portion of the SIN3 220 isoform designated the Uni-C consists of the unique stretch of 315 amino acids at the C-terminus of the SIN3 protein.

Genes encoding SIN3 complex proteins needed to be cloned into bacterial expression vectors to express them in *E. coli*. The pMAL-C2X, pET28a and pETLX vectors were used for the expression of recombinant proteins with affinity tags. The clones and their affinity tags are listed in Table 3.

Table 3. Bacterial Expression Clones

PROTEIN	VECTOR	TAG
SIN3 220	pMAL-C2X	MBP/HA[2]
Sap130	pETLX	HIS
SDS3	pET28a	HIS
ING1	pETLX	HIS
RPD3	pET28a	HIS
EMSY	pETLX	HIS
SIN3 220 Uni-C[1]	pMAL-C2X	MBP/HA[2]
Caf1-55[1]	pET28a	HIS

[1]denotes bacterial clone created/donated by Nirmalya Saha

[2]SIN3 220 clones had an HA-tag added, which is not a part of the vector

After cloning the correct sequences into expression vectors, the bacterial colonies harboring the plasmid of interest needed to be induced to express the proteins. The two approaches used to induce the expression of these proteins were IPTG induction and Studier Autoinduction. Both approaches use lactose or IPTG, a lactose analogue, to induce expression of the gene of interest. Expression of genes in the pET28a/pETLX vectors are under control of the T7 promoter while genes in the pMAL-c2X vector are regulated by the P-lac promoter.

The expression system of recombinant proteins in *E. coli* takes advantage of the lac operon. BL21(DE3) cells have an endogenous T7 polymerase gene under control of a *lac*UV5 promoter (Studier and Moffatt, 1986). This T7 gene is regulated by the lac repressor protein (LacI). In the absence of lactose, LacI will bind to the *lac*UV5 promoter in front of the gene encoding T7 polymerase. This prevents the *E. coli* RNA polymerase from transcribing the T7 polymerase gene, as the LacI protein prohibits the binding of

RNA polymerase to the promoter (Daber et al., 2007). When lactose is introduced to a bacterial culture, it will bind to LacI, which induces a conformational change in the protein, releasing it from the lac operator (Bell and Lewis, 2000; Daber et al., 2007). This allows the *E. coli* RNA polymerase to transcribe the T7 polymerase. The T7 polymerase can then bind to the T7 promoter located on the vector transformed into the bacteria and transcribe the recombinant protein of interest (Studier and Moffatt, 1986).

After the transcription and consequent expression of the gene of interest, the solubility of the recombinant protein must be determined. If folded correctly, the recombinant protein should be found in the soluble fraction of bacterial lysates. There are conditions, however, under which a protein will not fold correctly and will subsequently be found in the insoluble pellet, which are known as aggregates. Some factors that can lead to aggregation of recombinant proteins include temperature, amount of induction and the sequence of the protein (Baneyx and Mujacic, 2004). Faster rates of bacterial growth (caused by high temperature) or large amounts of induction (caused by a high concentration of IPTG or lactose) can be detrimental to protein production if the cellular processes are happening too rapidly to allow proper folding of proteins. Two potential solutions to these problems include lowering the temperature at which the culture is growing or inducing with a lower concentration of IPTG or lactose. There are certain scenarios, however, where the amino acid composition of the protein will not allow proper folding. In this case expression of a protein containing an amino acid tag to increase solubility (such as GST or MBP) can be utilized to express proteins in a soluble form (Esposito and Chatterjee, 2006; Raran-Kurussi and Waugh, 2012; Zhou and Wagner, 2010).

IPTG induction of recombinant protein expression is a common technique used in bacterial expression. There are times where induction via addition of IPTG, which is introduced when the culture has reached mid-log growth phase, results in insoluble proteins despite modifications of the amount of IPTG or temperature at which the culture is grown. One method to combat this problem is the use of Studier Autoinduction (Studier, 2005). While IPTG is added to a culture grown in LB in IPTG induction, autoinduction uses media that contains glucose, lactose, glycerol and salts. Instead of adding exogenous IPTG to the culture to induce expression of the recombinant protein, autoinduction media induces expression using components already present in the media. The autoinduction protocol takes advantage of the natural machinery of bacteria that first uses glucose as an energy source before switching to lactose. In turn, this is a more gradual shift towards protein expression. When the culture uses up the available glucose, the cells are in the stationary phase of their growth curve. At this point, in the majority of cells, there are more chaperone proteins available to help with protein folding, including the folding of the recombinant protein. Thus, autoinduction can provide an alternative method to produce a higher percentage of soluble protein than IPTG induction. For my studies, I tested the solubility of proteins expressed in bacteria using both approaches.

Before conducting immunoprecipitation experiments using bacterial lysates, it is beneficial to use previously solved structural data to hypothesize possible interactions between any two proteins. Although no structural studies of the entire SIN3 protein are available, there are some solved structures of conserved domains of the SIN3 protein, including the histone deacetylase interacting domain (HID) and paired amphipathic helices (PAHs). The Uni-C region of the SIN3 220 isoform, however, has no solved

structure and has been assayed for interactions with only one protein, Caf1-55 (Saha, 2017).

The amino acid sequence representing the unique C-terminus of the SIN3 220 isoform was analyzed using ANCHOR and IUPRED structural prediction software (Dosztányi et al., 2005a; Dosztányi et al., 2005b; Dosztányi et al., 2009). IUPRED software will estimate the energies of the potential folded vs. unfolded states of a stretch of amino acids to predict structures vs. unstructured regions (Dosztányi et al., 2005a; Dosztányi et al., 2005b). The IUPRED software incorporates an energy matrix derived from previous solved structures of globular proteins. The energy matrix predicts the energy of interactions between amino acids in a submitted peptide sequence. The matrix accounts for both the chemical properties as well as the immediate environment of the amino acid. If the energy generated by interactions is enough to overcome the energy lost during the folding of a stretch of amino acids then the amino acids would be predicted to be ordered. On the other hand, if the potential interactions are not energetically favorable, the given stretch of amino acids would be predicted to be disordered (Dosztányi et al., 2005b).

ANCHOR software aims to identify potential regions of a stretch of amino acids that may become more energetically favorable upon binding a globular protein (Mészáros et al., 2009). To do this, the analysis focuses on three criteria. First, the algorithm incorporates the generated IUPRED data to eliminate regions that would be prone to assuming a globular conformation. This allows for the identification of stretches of amino acids that are predicted to be unstructured. Second, the algorithm looks for regions of the input sequence that lack the energetic potential to interact with amino acids in their

immediate surrounding environment, thus resulting in a more unstructured region. Third, the algorithm analyzes the energetic favorability of potential interactions between globular proteins and amino acids of the input sequence. The software computes all three factors individually and then combines them to come up with a single predicted score, which estimates the potential of a stretch of amino acids interacting with a global protein (Dosztányi et al., 2009).

The sequence of the unique C-terminus of SIN3 220 was predicted to be unstructured when tested using IUPRED software (Figure 6). Any points plotted above the line at the 0.5 mark on the y-axis of the graph are predicted to be unstructured, while points that fall below the line are predicted to be structured. There were, however, regions of the unique C-terminus that were predicted by ANCHOR to have potential binding affinity for other proteins (Figure 6). Similarly, points that fall above the line at the 0.5 mark on the y-axis are predicted to have binding affinity for globular proteins, while points below the mark are not predicted to interact with another protein. It is important to note, however, that these are predictions and do not eliminate the need for *in vitro* and *in vivo* testing to determine structure and binding.

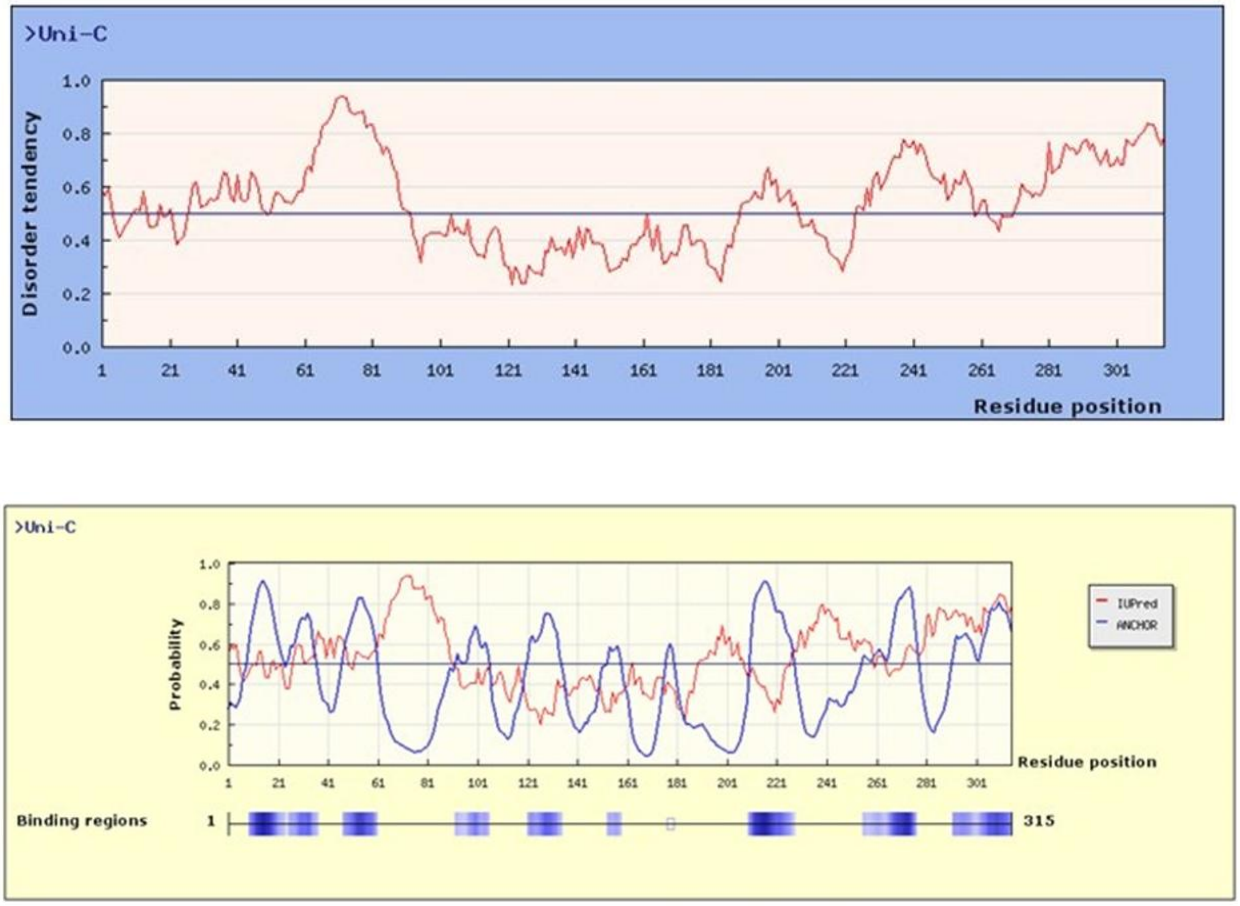

Figure 6. The unique C-terminus of SIN3 220 is disordered. The sequence representing the unique 315 amino acids at the C-terminus of SIN3 220 was analyzed using IUPRED software (red lines) and regions predicted to be disordered or ordered were identified. The software did not identify any regions predicted to be ordered (top). The sequence was also analyzed using ANCHOR software (blue lines), which looks for potential regions where binding to other proteins may occur (bottom). Potential binding regions are indicated by the blue boxes.

There is therefore a gap of knowledge of potential interacting partners of the Uni-C region (with the exception of Caf1-55), which if filled could lend insight into complex stability or the targeting of the complex to chromatin. To identify more potential *in vivo* binding partners of the C-terminus of SIN3 220, individual components of the SIN3 220

complex were cloned into bacterial expression vectors and immunoprecipitated along with the Uni-C region of 220.

3.2 Results and Discussion

Both IPTG and autoinduction approaches were used to express recombinant protein. Figure 7 shows examples of protein expressed with both the IPTG and autoinduction methods. Compared to the IPTG-induced proteins, proteins expressed via autoinduction media had a greater amount of protein found in the soluble fraction.

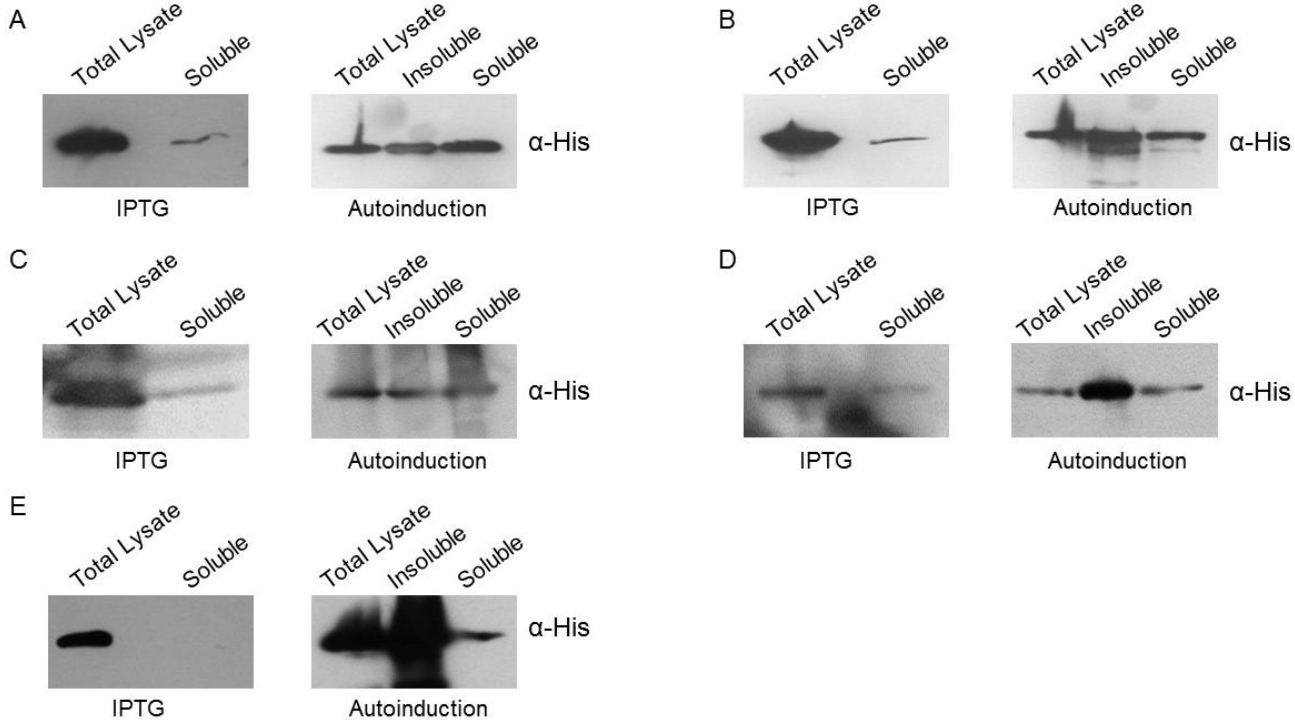

Figure 7. Comparison of IPTG and autoinduction. Bacterial cells were used to express recombinant proteins using either IPTG induction or autoinduction (Studier 2005). Cells were lysed and the extract subjected to fractionation. Fractions were analyzed by Western blot for ING1 (A), Sap130 (B), RPD3 (C), EMSY (D) and SDS3 (E).

Along with increased amounts of recombinant protein solubility, the culture grown using the autoinduction method reaches a higher density than cultures grown and induced with IPTG (data not shown). Since the autoinduction cultures are grown for a longer amount of time than IPTG-induced cultures, the density is increased, resulting in a larger number of cells available for fractionation and subsequent immunoprecipitation experiments. Therefore, due to the increased amount of soluble protein and denser cultures using autoinduction conditions as opposed to induction via IPTG, cell lysates were prepared by the autoinduction method for the immunoprecipitation experiments.

Immunoprecipitation experiments were then performed to assay the possible interaction between SIN3 220 Uni-C and other SIN3 complex proteins. The interaction between Caf1-55 and the Uni-C region of SIN3 220 was used as a positive control as this interaction has been previously seen using the bacterial expression system (Saha, 2017). An immunoprecipitation was performed using anti-HA resin and bacterial lysates containing HA-tagged SIN3 220 Uni-C and HIS-tagged Caf1-55. The bound fraction of the immunoprecipitation contained both SIN3 220 Uni-C and Caf1-55, which indicates a direct interaction between the two proteins (Figure 8A).

After confirming the validity of the system through the interaction of Caf1-55 and Uni-C, the interaction between the Uni-C region of SIN3 220 and EMSY was probed. As EMSY associates with the SIN3 220 isoform but not the 187 isoform (Spain et al., 2010), it was hypothesized that the Uni-C region of SIN3 220 may be responsible for interacting with EMSY and mediating the association with the SIN3 220 complex. It was found, however that the Uni-C region of SIN3 220 does not directly interact with EMSY when the bound fraction was analyzed by Western blot (Figure 8B).

Potential interactions between the Uni-C region of SIN3 220 and RPD3 (Figure 8D), SDS3 (Figure 8C) and Sap130 (Figure 8E) were then assayed. Upon analysis of the immunoprecipitated proteins through Western blot, it was determined that RPD3, SDS3 and Sap130 do not have affinity for the Uni-C region when expressed in a bacterial system (Figure 8). This result is not surprising, as all three proteins have either been shown or have been predicted to interact with regions present in both the 187 and 220 isoforms of the SIN3 protein. RPD3 interacts with the HID of SIN3, which is a domain that is highly conserved from yeast to mammals (Laherty et al., 1997). Additionally, the HID has been shown through structural studies to interact with SDS3 (Clark et al., 2015), and has been proposed to mediate interactions with Sap130 (Fleischer et al., 2003).

Interestingly, upon assaying the interaction between the Uni-C region of SIN3 220 and ING1, which is a protein common to both the 187 and 220 complexes (Fleischer et al., 2003), an interaction was observed using the bacterial expression system (Figure 8F). As a control, the immunoprecipitation was performed with ING1 in the absence of the HA-tagged Uni-C region, and no nonspecific binding was observed (Figure 8F). This finding indicates that the interaction between the Uni-C region of SIN3 220 and ING1 is not due to nonspecific binding of ING1 to the HA-resin. To the best of our knowledge this physical interaction between ING1 and the Uni-C region of SIN3 220 has not been previously demonstrated.

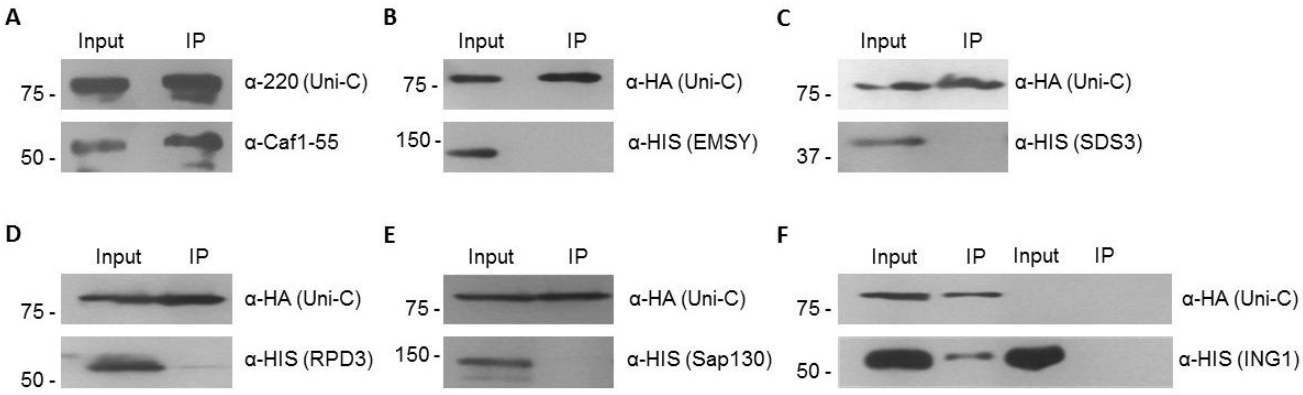

Figure 8. Interactions with SIN3 220 Uni-C. Recombinant SIN3 complex proteins were expressed using the autoinduction system for *E. coli* and co-immunoprecipitation experiments were performed to assay interactions with the Uni-C region of SIN3 220. All experiments were performed using anti-HA resin. The interaction of Caf1-55 and SIN3 220 Uni-C (A) was used as a positive control. No interactions were observed between Uni-C and EMSY (B), SDS3 (C), RPD3 (D) or Sap130 (E); an interaction was observed, however, between Uni-C and ING1 (F). Input and bound samples were analyzed by Western blot. IP – immunoprecipitation. Molecular weight size markers are indicated at the left of each blot.

It is possible that the unstructured form of the Uni-C region could act as a “sticky” peptide and could have affinity for a large amount of proteins and that those interactions have little to no biological relevance. As mentioned before, the C-terminus of SIN3 220 is not predicted to have any secondary structure. The lack of secondary structure could result in more lax requirements for the binding of other proteins. If the Uni-C region of SIN3 220 was sticky, it would be expected many more binding partners would have been co-immunoprecipitated with this protein. However, as there was no interaction between Uni-C and RPD3, SDS3 and Sap130, it is not likely that this unstructured protein is sticky. The lack of interaction between HIS-tagged ING1 and anti-HA resin strongly suggests that the positive interaction between ING1 and the Uni-C is genuine.

The biological significance of the possible interaction between ING1 and the unique C-terminus of SIN3 220 needs to be probed further. ING1 has intrinsic H3K4me3 binding ability mediated through its C-terminal plant homeodomain (PHD) finger, and this PHD finger binds specifically to methylated H3K4 (Peña et al., 2009). H3K4me3 is known to be a mark of active transcription (Barski et al., 2007). The association of ING1 and SIN3 and ING1's binding to H3K4me3 are consistent with the ability of SIN3 to act as a transcriptional activator (Gajan et al., 2016).

Additionally, ING1 has been demonstrated to be important for the recruitment of SIN3 to chromatin. Loss of ING1 in mammalian cells showed a reduction in binding of the SIN3 complex to the *pl2* promoter in mammalian cells (Smith et al., 2010). As SIN3 has been shown to regulate a significant portion of the genome, it is possible that large subsets of genes may be affected upon knockdown of *Ing1*. It would be interesting to compare the expression of genes between *Sin3a* and *Ing1* knockdown cells.

Both SIN3 and ING1 have been implicated in the development of cancer (Farias et al., 2010; Kwon et al., 2015; Thakur et al., 2014). Reduced levels of ING1 have been shown to result in breast cancer, and overexpression of ING1 can actually slow the rate of growth of cancerous cells (Thakur et al., 2014). Aberrant gene expression leading to cancer due to inappropriate levels of ING1 has led studies to look at the targeting of ING1 as a potential treatment for cancer patients (Zhang et al., 2017). The roles of both ING1 and SIN3 in the development of cancer could potentially be linked through their interaction.

Chapter 4 Analysis of the Interaction between Caf1-55 and SIN3

4.1 Introduction

Both Caf1-55 and SIN3 have been identified in previous work to be important for the regulation of genes in S2 cells (Saha, 2017). Additionally, phenotypes resulting from knockdown of *Caf1-55* mirror the changes in gene expression in S2 cells caused by the loss of SIN3 220 (Saha, 2017). Caf1-55 has been shown to interact with SIN3 *in vivo* through co-immunoprecipitation of tagged Caf1-55 and the C-terminus of SIN3 220 expressed in S2 cells (Saha, 2017). Caf1-55 has been shown to interact with the Uni-C region of SIN3 220 using bacterial expression followed by subsequent immunoprecipitation (Saha, 2017).

Despite identifying a direct interaction between SIN3 220 and Caf1-55, the region of Caf1-55 responsible for mediating this interaction was unknown. Previous work had looked to determine domains responsible for the interaction, and specifically investigated the histone H3 and histone H4 binding domains of Caf1-55 (Saha, 2017). Upon mutating select residues in the histone H3 binding domain of Caf1-55, no loss of SIN3 binding to Caf1-55 was observed. There was, however, a partial loss of SIN3 association with the histone H4 binding mutant of Caf1-55.

Studies have identified a truncation mutant of Caf1-55 (Caf1-55med), which leads to developmental defects in *Drosophila* as well as a reduction in the number of expected eclosed adults (Anderson et al., 2011). Furthermore, this mutation is able to act in a dominant negative fashion (Anderson et al., 2011). Even when wildtype Caf1-55 was expressed using the UAS/Gal-4 system, flies expressing the Caf1-55med mutant survived

to become pharate, but not many eclosed. Additionally, flies expressing Caf1-55med in the presence of UAS/Gal-4 induced wildtype Caf1-55 expression were found to have defects in eye shape, leg structures and wing defects when raised at 18^{o}C. Caf1-55med has a stop codon in place of the 304th amino acid, which is tryptophan. Caf1-55 is a protein that contains seven WD40 repeat domains. The Caf1-55med truncation results in a Caf1-55 protein that does not have the sixth and seventh WD40 repeat domains found in the wildtype protein. WD40 repeat domains create a surface that is used in protein-protein interactions (Suganuma et al., 2008). Furthermore, the region found to be responsible for facilitating the interaction between Caf1-55 and histone H4 is in the portion of the protein that is deleted in the Caf1-55med mutant (Song et al., 2008). Combining the knowledge of the role of Caf1-55 in regulating essential genes as well as the reduction of interaction between SIN3 and the histone H4 binding mutant of Caf1-55, we hypothesized that the interaction between SIN3 and Caf1-55 would be completely abolished in the presence of Caf1-55med.

4.2 Results and Discussion

Stable S2 cell lines expressing HA and FLAG-tagged Caf1-55 and Caf1-55med were used to assay the interactions between Caf1-55 and SIN3. Wildtype Caf1-55 was used as a positive control, as the interaction with SIN3 has already been observed (Saha, 2017). S2 cells carrying no transgene were used as a negative control. These control cells were treated in the same way as the transfected cell lines.

To confirm expression of the HA-tagged proteins, whole cell lysates were made from the transfected and wildtype non-transfected lines and were subjected to SDS-PAGE. The lysates were then analyzed by Western blotting using an antibody to HA.

Interestingly, Caf1-55med was found to have a lower level of expression (~3 fold) when compared to Caf1-55 (Figure 9). It is possible that the deleterious effects of Caf1-55med in *Drosophila* (Anderson et al., 2011) are also reflected in cell culture, and that the cell works to rid itself of the harmful protein through a degradation pathway.

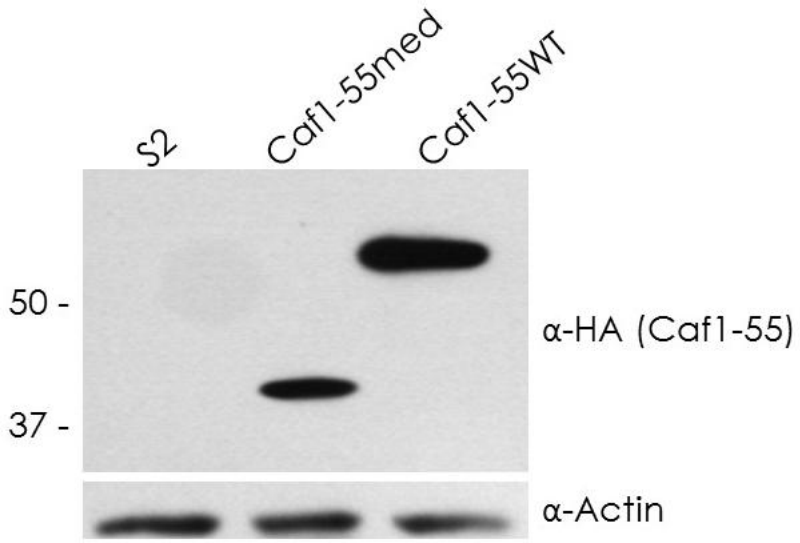

Figure 9. Expression of tagged Caf1-55. Transfected S2 cell lines for stable expression of HA-tagged Caf1-55WT or Caf1-55med were generated. Cells were induced with $CuSO_4$. Whole cell lysates were then subjected to SDS-PAGE and proteins visualized by Western blot using antibody against the HA-tag of the transgenic proteins. An S2 cell line carrying no transgene was used as a negative control. Actin was used as a loading control.

Nuclear fractions were prepared from the transfected and non-transfected lines and used for immunoprecipitation. When assaying the interaction between the Caf1-55 proteins and SIN3, different amounts of the bound fractions were subjected to SDS-PAGE to account for the differences in expression between the Caf1-55 and Caf1-55med tagged

proteins. This ensured an approximately equivalent amount of protein in each bound fraction was analyzed. Fractions eluted from the resin were subjected to SDS-PAGE and Western blotting for analysis. The interaction between Caf1-55 and SIN3 was observed; however, there was no observable interaction between Caf1-55med and SIN3 (Figure 10). It is therefore likely that SIN3 interacts with Caf1-55 through the sixth or seventh WD40 repeat region of Caf1-55.

The co-immunoprecipitation of Caf1-55 and the unique C-terminus of SIN3 using the bacterial expression system found that the two proteins are interacting partners. Additionally, the interaction of Caf1-55 and SIN3 was not observed in S2 cells when the last two WD40 repeat domains of Caf1-55 were deleted. These two pieces of data lend possible insight into the interacting regions of these two proteins. It is entirely possible that the Uni-C region of SIN3 has affinity for one or more WD40 repeat domains at the C-terminus of Caf1-55, and that these regions are responsible for mediating the interaction between the two proteins. More studies would have to be performed in order to determine the exact interacting regions, such as mutating select residues or deleting smaller regions of the proteins and then assaying the interaction.

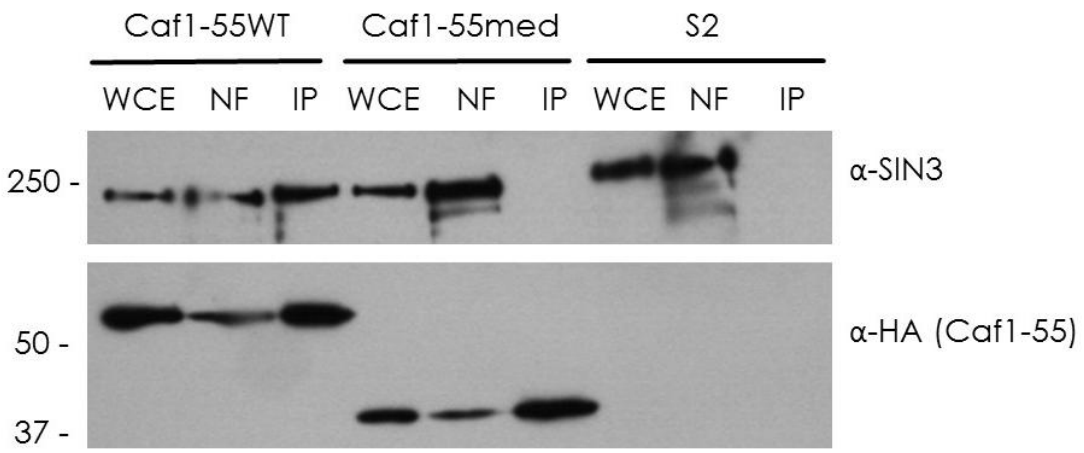

Figure 10. Loss of interaction between SIN3 and Caf1-55med. Immunoprecipitation experiments were performed against HA-tagged Caf1-55WT and Caf1-55med. Wildtype S2 cells were used as a negative control. Fractions were subjected to Western blotting and were probed with antibody against SIN3 and HA as indicated. WCE - Whole Cell Lysate, NF- Nuclear Fraction, IP – Immunoprecipitation. Molecular weight size markers are indicated at the left of each blot.

It would also be of interest to look at gene expression changes in cells expressing Caf1-55med. As previously mentioned, a number of genes have been found to change in expression when both *Sin3a* and *Caf1-55* are knocked down in S2 cells (Saha, 2017; Saha et al., 2016). Disrupting the interaction between SIN3 and Caf1-55 may lead to similar changes in gene expression. The data shown above concerning differential expression of Caf1-55 and Caf1-55med have indicated that S2 cells may attempt to alleviate the deleterious effects of Caf1-55med by targeting it for degradation. This differential level of expression is in line with the results that a mutation leading to expression of Caf1-55med instead of the full-length Caf1-55 can act in a dominant negative fashion (Anderson et al., 2011). It is possible that no changes in gene expression will be observed if the cell still has endogenous Caf1-55 to use. It would therefore be

preferential if endogenous Caf1-55 was knocked down in the context of expression of Caf1-55med.

Chapter 5 Summary

This study has worked to define interacting proteins within the *Drosophila* SIN3 complex. The regulation of a cell's genome is vital for maintaining the health of a cell and, consequently, the whole organism (Lee and Young, 2013; Villard, 2004; Wang et al., 2016; Wen et al., 2016). For example, the disruption of protein interactions can lead to destabilized protein complexes (Gibson, 2009; Perkins et al., 2010). In turn, this disruption can result in harmful consequences to the cell (Sato et al., 2000; Sherman et al., 2001).

The *Drosophila* SIN3 220 and 187 isoforms are capable of acting as a scaffold for two protein complexes that share the same core proteins, with the 220 isoform additionally interacting with Caf1-55, LID and EMSY (Spain et al., 2010). As these two isoforms regulate similar yet distinct genes (Saha et al., 2016), it was of interest to identify SIN3 complex proteins that are capable of interacting with the unique 315 amino acid C-terminal region of the SIN3 220 isoforms. Caf1-55 had previously been identified as an interacting partner with the unique C-terminus of SIN3 220 (Saha, 2017). Using proteins generated with the bacterial expression system, it was found that EMSY does not interact with the unique C-terminus of SIN3 220, which leads us to hypothesize that EMSY may directly interact with Caf1-55 or LID, the other two unique components of the SIN3 220 complex.

Interestingly, ING1 was found to directly interact with the unique C-terminus of SIN3 220, despite being found in both the 220 and 187 complexes (Spain et al., 2010). As ING1 contains a PHD which binds to H3K4me3 (Peña et al., 2009), it would be of interest to see the effects that knockdown of *Ing1* has on the expression of genes regulated by SIN3 220. It is hypothesized that the interaction between ING1 and the

unique 315 amino acid C-terminus of SIN3 220 may be important for stability of the complex, which in turn may impact gene regulation. The data generated in this study and data concerning protein interactions from previous studies has allowed for a more thorough model concerning the protein interactions within the SIN3 complex (Figure 11).

Additionally, a region of Caf1-55 responsible for maintaining interaction with SIN3 was determined. A truncated version of Caf1-55 (Caf1-55med) previously identified and known to cause developmental defects in *Drosophila* (Anderson et al., 2011) was used for immunoprecipitation in S2 cells. When Caf1-55med was immunoprecipitated and the bound fraction was probed for the presence of SIN3, no association with SIN3 was observed. These data, along with the data from the bacterial expression system, allows us to hypothesize that the region deleted in Caf1-55med and the unique C-terminus of SIN3 220 are responsible for mediating their interaction. Future experiments will look at the gene expression of genes regulated by SIN3 in the presence of Caf1-55med.

The other known direct interactions taking place within the complex could lead to potential treatments for different genetic diseases. As already mentioned, the disruption of the interaction between Pf1 and SIN3 partially reduces the ability of breast cancer tumor cells to form colonies (Bansal et al., 2015). It would be interesting to see if similar genes change in expression upon the disruption of the SIN3-Pf1 interaction compared to knockdown of SIN3 alone. Additionally, comparing gene expression changes between the loss of the SIN3-Pf1 interaction and upon knockdown of Caf1-55 may be interesting, as loss of Caf1-55 resulted in gene expression changes similar to those seen upon loss of SIN3 (Saha, 2017).

Sap130 has been shown to be capable of repressing transcription. Interestingly, Sap130 is also hypothesized to interact with a transcriptional activator. (Fleischer et al., 2003). Reduction of Sap130 by RNAi in *Drosophila* can lead to a reduction of SIN3 at chromatin (Laity et al., unpublished). Additionally, mutations in Sap130 have demonstrated that Sap130 may be involved in mediating Aβ toxicity through chromatin regulation and assembly (Cao et al., 2008). These data combined with the interaction with SIN3 and ING1 provide rationale for future studies of Sap130, including analyzing gene expression and interaction domains.

Unlike Sap130, knockdown of SDS3 does not result in a loss of SIN3 at chromatin (Laity et al. unpublished). Similar to Sap130, however, SDS3 has been shown to directly interact with SIN3 and have a role in repressing transcription (Cao et al., 2008). Moreover, loss of SDS3 has been shown to result in embryonic lethality in mammals , and mutants of SDS3 unable to bind to SIN3 are not able to rescue the phenotype (David et al., 2003). The function of SDS3 is not consistent across organisms, however, as studies in mammalian cells have conflicted with findings in yeast (Clark et al., 2015). While yeast require SDS3 to recruit HDAC1 to SIN3 complexes, HDAC1 is recruited to SIN3 complexes through a direct interaction with SIN3 in mammals. Studies involving SDS3 in *Drosophila* may result in new and novel approach to understand the molecular basis by which SDS3 functions and the role it plays in the regulation of genetic material.

The results of this study have provided a basis for numerous questions and experiments that will further probe the roles of protein interactions within the *Drosophila* SIN3 complex. Combining my data with previously published data, a model that more accurately reflects the spatial arrangement of proteins within the SIN3 complex has been

generated (Figure 11). As SIN3 and its associated proteins have been shown to play an important role in gene regulation, further studies could lend potential insights into treatments for various genetic diseases, including cancer.

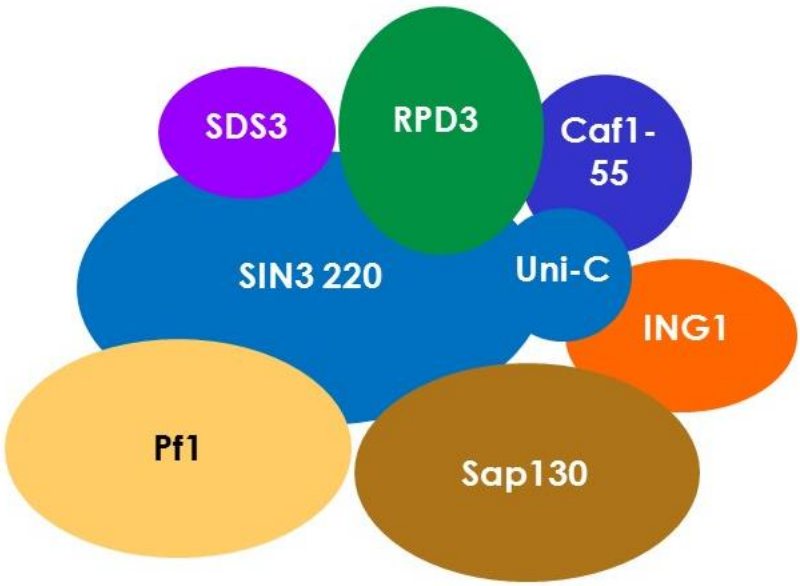

Figure 11. Currently known direct SIN3 complex protein interactions. Data from this research and previous work (Figure 3) has helped create a more detailed understanding of the interactions within the *Drosophila* SIN3 complex. The model shows all known protein interactions to data, with the unique C-terminus of the SIN3 220 isoform shown as “Uni-C”.

References

Anderson, A.E., Karandikar, U.C., Pepple, K.L., Chen, Z., Bergmann, A., and Mardon, G. (2011). The enhancer of trithorax and polycomb gene Caf1/p55 is essential for cell survival and patterning in Drosophila development. Development *138*, 1957-1966.

Aregger, M., and Cowling, V.H. (2016). Regulation of mRNA capping in the cell cycle. RNA Biol, 0.

Baneyx, F., and Mujacic, M. (2004). Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol *22*, 1399-1408.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381-395.

Bansal, N., Petrie, K., Christova, R., Chung, C.Y., Leibovitch, B.A., Howell, L., Gil, V., Sbirkov, Y., Lee, E., Wexler, J.*, et al.* (2015). Targeting the SIN3A-PF1 interaction inhibits epithelial to mesenchymal transition and maintenance of a stem cell phenotype in triple negative breast cancer. Oncotarget *6*, 34087-34105.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bell, C.E., and Lewis, M. (2000). A closer view of the conformation of the Lac repressor bound to operator. Nat Struct Biol *7*, 209-214.

Berger, S.L. (2002). Histone modifications in transcriptional regulation. Curr Opin Genet Dev *12*, 142-148.

Cao, W., Song, H.J., Gangi, T., Kelkar, A., Antani, I., Garza, D., and Konsolaki, M. (2008). Identification of novel genes that modify phenotypes induced by Alzheimer's beta-amyloid overexpression in Drosophila. Genetics *178*, 1457-1471.

Chaubal, A., Todi, S.V., and Pile, L.A. (2016). Inter-isoform-dependent Regulation of the Drosophila Master Transcriptional Regulator SIN3. J Biol Chem *291*, 11566-11571.

Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., Huang, L., Wen, Z., Li, W., Li, X.*, et al.* (2013). H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. Genes Dev *27*, 2109-2124.

Clark, M.D., Marcum, R., Graveline, R., Chan, C.W., Xie, T., Chen, Z., Ding, Y., Zhang, Y., Mondragón, A., David, G.*, et al.* (2015). Structural insights into the assembly of the histone deacetylase-associated Sin3L/Rpd3L corepressor complex. Proc Natl Acad Sci U S A *112*, E3669-3678.

Cowley, S.M., Iritani, B.M., Mendrysa, S.M., Xu, T., Cheng, P.F., Yada, J., Liggitt, H.D., and Eisenman, R.N. (2005). The mSin3A chromatin-modifying complex is essential for embryogenesis and T-cell development. Mol Cell Biol *25*, 6990-7004.

Daber, R., Stayrook, S., Rosenberg, A., and Lewis, M. (2007). Structural analysis of lac repressor bound to allosteric effectors. J Mol Biol *370*, 609-619.

Dannenberg, J.H., David, G., Zhong, S., van der Torre, J., Wong, W.H., and Depinho, R.A. (2005). mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. Genes Dev *19*, 1581-1595.

David, G., Turner, G.M., Yao, Y., Protopopov, A., and DePinho, R.A. (2003). mSin3-associated protein, mSds3, is essential for pericentric heterochromatin formation and chromosome segregation in mammalian cells. Genes Dev *17*, 2396-2405.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491-496.

Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005a). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433-3434.

Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I. (2005b). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol *347*, 827-839.

Dosztányi, Z., Mészáros, B., and Simon, I. (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics *25*, 2745-2746.

Esposito, D., and Chatterjee, D.K. (2006). Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotechnol *17*, 353-358.

Farias, E.F., Petrie, K., Leibovitch, B., Murtagh, J., Chornet, M.B., Schenk, T., Zelent, A., and Waxman, S. (2010). Interference with Sin3 function induces epigenetic reprogramming and differentiation in breast cancer cells. Proc Natl Acad Sci U S A *107*, 11811-11816.

Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet *13*, 97-109.

Fleischer, T.C., Yun, U.J., and Ayer, D.E. (2003). Identification and characterization of three new components of the mSin3A corepressor complex. Mol Cell Biol *23*, 3456-3467.

Gajan, A., Barnes, V.L., Liu, M., Saha, N., and Pile, L.A. (2016). The histone demethylase dKDM5/LID interacts with the SIN3 histone deacetylase complex and shares functional similarities with SIN3. Epigenetics Chromatin *9*, 4.

Gan, H., Hao, Q., Idell, S., and Tang, H. (2016). Interferon-γ Promotes Double-Stranded RNA-Induced TLR3-Dependent Apoptosis via Upregulation of Transcription Factor Runx3 in Airway Epithelial Cells. Am J Physiol Lung Cell Mol Physiol, ajplung.00278.02016.

Garcia-Sanz, P., Quintanilla, A., Lafita, M.C., Moreno-Bueno, G., García-Gutierrez, L., Tabor, V., Varela, I., Shiio, Y., Larsson, L.G., Portillo, F.*, et al.* (2014). Sin3b interacts with Myc and decreases Myc levels. J Biol Chem *289*, 22221-22236.

Gibson, T.J. (2009). Cell regulation: determined to signal discrete cooperation. Trends Biochem Sci *34*, 471-482.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E.*, et al.* (2003). A protein interaction map of Drosophila melanogaster. Science *302*, 1727-1736.

Hiragami-Hamada, K., Soeroes, S., Nikolov, M., Wilkins, B., Kreuz, S., Chen, C., De La Rosa-Velázquez, I.A., Zenn, H.M., Kost, N., Pohl, W.*, et al.* (2016). Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin. Nat Commun *7*, 11310.

Jackson, R.L., Busch, S.J., and Cardin, A.D. (1991). Glycosaminoglycans: molecular properties, protein interactions, and role in physiological processes. Physiol Rev *71*, 481-539.

Kadamb, R., Mittal, S., Bansal, N., Batra, H., and Saluja, D. (2013). Sin3: insight into its transcription regulatory functions. Eur J Cell Biol *92*, 237-246.

Kunowska, N., Rotival, M., Yu, L., Choudhary, J., and Dillon, N. (2015). Identification of protein complexes that bind to histone H3 combinatorial modifications using super-SILAC and weighted correlation network analysis. Nucleic Acids Res *43*, 1418-1432.

Kwon, Y.J., Petrie, K., Leibovitch, B.A., Zeng, L., Mezei, M., Howell, L., Gil, V., Christova, R., Bansal, N., Yang, S.*, et al.* (2015). Selective Inhibition of SIN3 Corepressor with Avermectins as a Novel Therapeutic Strategy in Triple-Negative Breast Cancer. Mol Cancer Ther *14*, 1824-1836.

Laherty, C.D., Yang, W.M., Sun, J.M., Davie, J.R., Seto, E., and Eisenman, R.N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. Cell *89*, 349-356.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73-84.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237-1251.

Letovsky, S., and Kasif, S. (2003). Predicting protein function from protein/protein interaction data: a probabilistic approach. Bioinformatics *19 Suppl 1*, i197-204.

Liang, M., Cornell, H.R., Zargar Baboldashti, N., Thompson, M.S., Carr, A.J., and Hulley, P.A. (2012). Regulation of hypoxia-induced cell death in human tenocytes. Adv Orthop *2012*, 984950.

Marhold, J., Brehm, A., and Kramer, K. (2004). The Drosophila methyl-DNA binding protein MBD2/3 interacts with the NuRD complex via p55 and MI-2. BMC Mol Biol *5*, 20.

Mészáros, B., Simon, I., and Dosztányi, Z. (2009). Prediction of protein binding regions in disordered proteins. PLoS Comput Biol *5*, e1000376.

Mishima, Y., Jayasinghe, C.D., Lu, K., Otani, J., Shirakawa, M., Kawakami, T., Kimura, H., Hojo, H., Carlton, P., Tajima, S.*, et al.* (2015). Nucleosome compaction facilitates HP1γ binding to methylated H3K9. Nucleic Acids Res *43*, 10200-10212.

Neufeld, T.P., Tang, A.H., and Rubin, G.M. (1998). A genetic screen to identify components of the sina signaling pathway in Drosophila eye development. Genetics *148*, 277-286.

Peña, P.V., Musselman, C.A., Kuo, A.J., Gozani, O., and Kutateladze, T.G. (2009). NMR assignments and histone specificity of the ING2 PHD finger. Magn Reson Chem *47*, 352-358.

Pennetta, G., and Pauli, D. (1998). The Drosophila Sin3 gene encodes a widely distributed transcription factor essential for embryonic viability. Dev Genes Evol *208*, 531-536.

Perkins, J.R., Diboun, I., Dessailly, B.H., Lees, J.G., and Orengo, C. (2010). Transient protein-protein interactions: structural, functional, and network properties. Structure *18*, 1233-1243.

Pile, L.A., Schlag, E.M., and Wassarman, D.A. (2002). The SIN3/RPD3 deacetylase complex is essential for G(2) phase cell cycle progression and regulation of SMRTER corepressor levels. Mol Cell Biol *22*, 4965-4976.

Pile, L.A., and Wassarman, D.A. (2000). Chromosomal localization links the SIN3-RPD3 complex to the regulation of chromatin condensation, histone acetylation and gene expression. The EMBO journal *19*, 6131-6140.

Raran-Kurussi, S., and Waugh, D.S. (2012). The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. PLoS One *7*, e49589.

Rossetto, D., Avvakumov, N., and Côté, J. (2012). Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics *7*, 1098-1108.

Saha, N. (2017). FUNCTIONAL ANAYSIS OF SIN3 ISOFORMS IN DROSOPHILA. In Department of Biological Sciences (Wayne State University).

Saha, N., Liu, M., Gajan, A., and Pile, L.A. (2016). Genome-wide studies reveal novel and distinct biological pathways regulated by SIN3 isoforms. BMC Genomics *17*, 111.

Sambrook J, R.J. (2001). Molecular cloning- a laboratory manual (CSHL Press).

Sato, T.K., Rehling, P., Peterson, M.R., and Emr, S.D. (2000). Class C Vps protein complex regulates vacuolar SNARE pairing and is required for vesicle docking/fusion. Mol Cell *6*, 661-671.

Schwikowski, B., Uetz, P., and Fields, S. (2000). A network of protein-protein interactions in yeast. Nat Biotechnol *18*, 1257-1261.

Sharma, V., Swaminathan, A., Bao, R., and Pile, L.A. (2008). Drosophila SIN3 is required at multiple stages of development. Dev Dyn *237*, 3040-3050.

Sherman, D.L., Fabrizi, C., Gillespie, C.S., and Brophy, P.J. (2001). Specific disruption of a schwann cell dystrophin-related protein complex in a demyelinating neuropathy. Neuron *30*, 677-687.
Smith, K.T., Martin-Brown, S.A., Florens, L., Washburn, M.P., and Workman, J.L. (2010). Deacetylase inhibitors dissociate the histone-targeting ING2 subunit from the Sin3 complex. Chem Biol *17*, 65-74.

Song, J.J., Garlick, J.D., and Kingston, R.E. (2008). Structural basis of histone H4 recognition by p55. Genes Dev *22*, 1313-1318.

Spain, M.M., Caruso, J.A., Swaminathan, A., and Pile, L.A. (2010). Drosophila SIN3 isoforms interact with distinct proteins and have unique biological functions. J Biol Chem *285*, 27457-27467.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif *41*, 207-234.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol *189*, 113-130.

Suganuma, T., Pattenden, S.G., and Workman, J.L. (2008). Diverse functions of WD40 repeat proteins in histone recognition. Genes Dev *22*, 1265-1268.

Swaminathan, A., and Pile, L.A. (2010). Regulation of cell proliferation and wing development by Drosophila SIN3 and String. Mech Dev *127*, 96-106.

Thakur, S., Singla, A.K., Chen, J., Tran, U., Yang, Y., Salazar, C., Magliocco, A., Klimowicz, A., Jirik, F., and Riabowol, K. (2014). Reduced ING1 levels in breast cancer promotes metastasis. Oncotarget *5*, 4244-4256.

Villard, J. (2004). Transcription regulation and human diseases. Swiss Med Wkly *134*, 571-579.

Volynets, G.P., Gorbatiuk, O.B., Kukharenko, O.P., Usenko, M.O., and Yarmoluk, S.M. (2016). Production of recombinant human apoptosis signal-regulating kinase 1 (ASK1) in Escherichia coli. Protein Expr Purif *126*, 89-92.

Wang, L., Ma, H., Zhu, L., Ma, L., Cao, L., Wei, H., and Xu, J. (2016). Screening for the optimal gene and functional gene sets related to breast cancer using differential co-expression and differential expression analysis. Cancer Biomark.

Weissman, I.L. (2015). Stem cells are units of natural selection for tissue formation, for germline development, and in cancer development. Proc Natl Acad Sci U S A *112*, 8922-8928.

Wen, X., Lu, R., Xie, S., Zheng, H., Wang, H., Wang, Y., Sun, J., Gao, X., and Guo, L. (2016). APE1 overexpression promotes the progression of ovarian cancer and serves as a potential therapeutic target. Cancer Biomark *17*, 313-322.
Yochum, G.S., and Ayer, D.E. (2001). Pf1, a novel PHD zinc finger protein that links the TLE corepressor to the mSin3A-histone deacetylase complex. Mol Cell Biol *21*, 4110-4118.

Zhang, C.J., Hou, X.M., Tan, L.M., Shao, C.R., Huang, H.W., Li, Y.Q., Li, L., Cai, T., Chen, S., and He, X.J. (2016). The Arabidopsis acetylated histone-binding protein BRAT1 forms a complex with BRP1 and prevents transcriptional silencing. Nat Commun *7*, 11715.

Zhang, R., Jin, J., Shi, J., and Hou, Y. (2017). INGs are potential drug targets for cancer. J Cancer Res Clin Oncol *143*, 189-197.

Zhang, W., Prakash, C., Sum, C., Gong, Y., Li, Y., Kwok, J.J., Thiessen, N., Pettersson, S., Jones, S.J., Knapp, S.*, et al.* (2012). Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells. J Biol Chem *287*, 43137-43155.

Zhou, P., and Wagner, G. (2010). Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies. J Biomol NMR *46*, 23-31.

Zilfou, J.T., Hoffman, W.H., Sank, M., George, D.L., and Murphy, M. (2001). The corepressor mSin3a interacts with the proline-rich domain of p53 and protects p53 from proteasome-mediated degradation. Mol Cell Biol *21*, 3974-3985.

ABSTRACT

DETERMINATION OF THE DIRECT PROTEIN-PROTEIN INTERACTIONS IN THE *DROSOPHILA* SIN3A COMPLEX

by

IAN MICHAEL MOORE

February 2017

Advisor: Lori A. Pile

Major: Biological Sciences

Degree: Master of Science

The proteins that comprise the *Drosophila* SIN3 220 and SIN3 187 complexes are currently known. Limited information with regard to interacting complex member proteins has been described. Much of these data are results of high-throughput investigations, and there have been no studies done to reconstitute all direct interactions within the complex. The unique C-terminal region of the SIN3 220 isoform was used to test for interactions with other complex proteins using the bacterial expression system. Additionally, the region of Caf1-55 necessary for interaction with SIN3 *in vivo* was identified using a truncation mutant. The results of this work identified novel protein interactions as well as novel regions of Caf1-55 required to maintain certain protein interactions *in vivo*. This study gives a clearer understanding of direct interacting partners within the *Drosophila* SIN3 complex as well as raising interesting questions regarding their role in gene regulation.

AUTOBIOGRAPHICAL STATEMENT

IAN MICHAEL MOORE

Raised in Michigan, I grew up in Southgate, a suburb of Detroit and attended Southgate Anderson High School.

I attended the University of Michigan-Dearborn for undergraduate studies, where I majored in Biological Sciences, and I minored in Anthropology.

Following undergrad, I attended Wayne State University for my Masters of Science in Biology, and am looking to start a career in scientific writing.